

# Использование сверточных нейронных сетей в задаче преобразования цветных координат<sup>1</sup>

А. С. Срибняк<sup>\*,\*\*</sup>, А. В. Паншин<sup>\*,\*\*</sup>, Е. И. Ершов<sup>\*,\*\*</sup>

<sup>\*</sup> Институт проблем передачи информации имени А.А. Харкевича, Российская академия наук, Москва, Россия

<sup>\*\*</sup> Московский физико-технический институт (национальный исследовательский университет), Москва, Россия

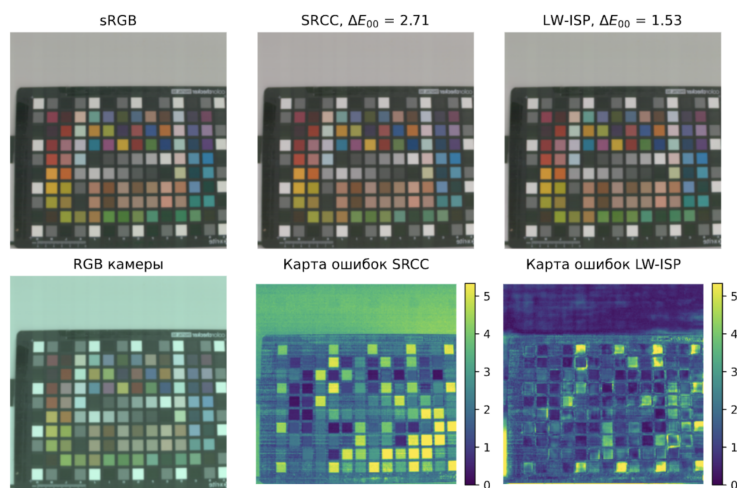
Поступила в редколлегию 01.07.2025 г. Принята 28.07.2025 г.

**Аннотация**—Данная работа посвящена исследованию задачи преобразования цветных координат — части конвейера формирования цифрового изображения. Существующие методы решения этой задачи в основном преобразуют изображение поточечно, без учета контекста. В работе представлен новый метод преобразования цветных координат на основе сверточной нейронной сети LW-ISP, способный учитывать семантику сцены. Предлагаемый метод демонстрирует значительное улучшение по сравнению с современными подходами к цветовым переходам: снижение цветового различия CIEDE2000 на 39%, и уменьшение углового расстояния в пространстве ProLab на 35%.

**КЛЮЧЕВЫЕ СЛОВА:** преобразование цветных координат, цветокоррекция, сверточные нейронные сети.

DOI: 10.53921/18195822\_2025\_25\_2\_224

## 1. ВВЕДЕНИЕ



**Рис. 1.** Иллюстрация работы предлагаемого нейросетевого метода CST, основанного на модели LW-ISP, в сравнении с SRCC — одним из наиболее точных из поточечных методов. На картах ошибок визуализирована метрика CIEDE2000 ( $\Delta E_{00}$ ).

<sup>1</sup> Исследование осуществлено в рамках Госзадания № FFNU-2022-0028 ИППИ РАН

В процессе первичной обработки изображения, снятого на цифровую камеру, решается целый ряд задач: подавление шума, демозаикинг, баланс белого, преобразование цветовых координат, гамма-коррекция и т.д. Этот процесс называется конвейером формирования изображения, в англоязычной литературе — ISP pipeline, Image Signal Processing pipeline. Он необходим для приведения необработанных данных с сенсора камеры к такому виду, который при выводе на экран или печати будет выглядеть достаточно хорошо с точки зрения пользователя или достаточно похоже на запечатленную сцену — конкретная цель зависит от области применения. Задача преобразования цветовых координат (CST, Color Space Transform) является одной из важнейших частей ISP. Она отвечает за передачу цветов на фотографии такими, какими их видит человек. Нетривиальность этой задачи вызвана тем, что у человеческого глаза и у камер цифровых фотоаппаратов или смартфонов значительно различаются функции спектральной чувствительности. Согласно критерию Максвелла-Лютера-Айвса [15], если не существует линейного преобразования между спектральными чувствительностями двух сенсоров, то не существует однозначного соответствия между откликами этих сенсоров. На практике чувствительности камеры практически никогда в точности не выражаются через чувствительности человеческого глаза, поэтому некоторые цвета, которые различны для человека, неразличимы для камеры и наоборот.

В качестве практического и эффективного решения широко применяется линейная цветокоррекция. Этот метод преобразует отклики линейно, умножая каждый пиксель изображения на матрицу  $3 \times 3$ . Ключевыми преимуществами линейного подхода являются его вычислительная эффективность и способность обеспечивать приемлемое качество цветопередачи в большинстве сцен. Оптимальная матрица обычно находится с помощью метода наименьших квадратов. Важное свойство линейной модели — однородность порядка 1: при умножении входного вектора на скаляр выходной вектор умножается на тот же скаляр. Физический смысл этого свойства заключается в том, что при изменении яркости в цветовом пространстве камеры соответствующим образом меняется яркость в цветовом пространстве человека без изменения тона или насыщенности. Естественный способ улучшения точности CST — добавить в линейную модель нелинейные признаки. Добавив полиномиальные признаки степени не больше  $d$ , получаем полиномиальную цветокоррекцию (PCC, Polynomial Color Correction) порядка  $d$ . Полиномиальная модель уже не обладает свойством однородности порядка 1, из-за чего качество реконструкции фотографии может сильно отличаться для разных уровней яркости. Чтобы обеспечить однородность модели одновременно с приростом качества, были предложены корневая полиномиальная цветокоррекция (RPCC, Root-Polynomial Color Correction) [3] и более вычислительно эффективная цветокоррекция с однородными рациональными признаками (SRCC, Scalable Rational Color Correction) [14].

В работе [12] было предложено использовать для цветокоррекции полносвязную нейронную сеть. Этот метод, как и описанные выше принимает на вход 3 числа и возвращает 3 числа — три стимулы в двух цветовых пространствах. Авторы предложили архитектуру сети с 3 скрытыми слоями из 21, 77 и 21 нейрона соответственно, с логистической функцией в качестве функции активации после каждого из них. Позже в статье [13] те же авторы предложили использовать сеть с 2 скрытыми слоями из 79 и 36 нейронов соответственно.

Несмотря на заявляемое высокое качество, использование полносвязной нейросети для CST имеет очевидные недостатки, среди которых вычислительная неэффективность (много операций, в том числе вычисление логистической функции) и невыполнение в общем случае свойства однородности, что было отмечено в работе [7]. В статье [8] те же авторы предлагают модификацию нейросетевого решения, обладающую свойством однородности.

Логичный следующий шаг в развитии методов CST — использование моделей, учитывающих не только цвет в одной точке изображения, но и пространственный контекст. В частно-

сти, к таким относятся сверточные нейронные сети. Тем не менее, на данный момент описано лишь применение сверточных нейросетей к смежным задачам, например, к задаче цветовой константности [1, 18] и цветокоррекции подводных изображений [16, 9, 4]. В этих задачах важность контекста очевидна, потому закономерно использование таких моделей. В данной работе мы демонстрируем, что и для задачи CST семантика сцены важна, решая эту задачу с помощью сверточной нейронной сети.

## 2. ПОСТАНОВКА ЗАДАЧИ ПРЕОБРАЗОВАНИЯ ЦВЕТОВЫХ КООРДИНАТ

Введем формальную постановку задачи преобразования цветковых координат. *Сенсором* будем называть часть цифровой камеры или зрительной системы человека, регистрирующую цветковые свойства достигшего системы излучения. Отклик сенсора при наблюдении некоторого объекта зависит от приходящего на сенсор спектра излучения  $F(\lambda)$  и функции спектральной чувствительности сенсора  $\chi(\lambda)$ :  $a = \int_0^\infty F(\lambda)\chi(\lambda)d\lambda$ , где  $\lambda > 0$  — длина волны излучения. В глазу человека спектр регистрируется 3 видами клеток-колбочек с различными спектральными чувствительностями:  $\vec{\chi}(\lambda) = (\chi_1(\lambda), \chi_2(\lambda), \chi_3(\lambda))^T$ . Вектор их откликов, называемый *тристимулом*, характеризует воспринимаемый цвет [21]:

$$\vec{v} = \int_0^\infty F(\lambda)\vec{\chi}(\lambda)d\lambda. \quad (1)$$

Аналогично в цифровых камерах преимущественно устанавливается по 3 вида фильтров, определяющих ее функции спектральной чувствительности, то есть модель цветовосприятия камеры описывается той же формулой. Тройку откликов также называют *цветовыми координатами*.

Заметим, что совершенно разные спектры при таком отображении могут переходить в одинаковые векторы откликов, то есть становиться неразличимыми для данного сенсора. Для фиксированного сенсора определяются понятия *цвета* как множества неразличимых излучений и *цветового пространства* как множества всех цветов. Цвет однозначно описывается цветковыми координатами.

Если два цвета имеют цветковые координаты, отличающимися умножением на скаляр ( $\vec{v}$  и  $a\vec{v}$ ), то говорят, что они имеют одинаковую *цветность*, но разную *яркость*. Не связанные таким соотношением цвета имеют разную цветность [19].

Включая в модель (1) зависимость от пространственных координат  $x, y \in D \subset R^2$ , получаем модель формирования цветного изображения:

$$I(x, y) = \int_0^\infty F(\lambda, x, y)\vec{\chi}(\lambda)d\lambda. \quad (2)$$

Если все объекты на изображении освещены одним источником освещения, или несколькими источниками с одинаковым спектральным составом излучения, можно также разложить  $F$  в произведение  $F(\lambda, x, y) = S(\lambda)\Phi(\lambda, x, y)$ , где  $S(\lambda)$  — спектр излучения источника,  $\Phi(\lambda, x, y)$  — отражательная характеристика сцены. Для такого случая модель формирования изображения переписывается в виде:

$$I(x, y) = \int_0^\infty S(\lambda)\Phi(\lambda, x, y)\vec{\chi}(\lambda)d\lambda. \quad (3)$$

Эту модель часто используют и в случае более сложного освещения как удобное упрощение.

Задача *преобразования цветковых координат* (CST, Color Space Transform) — по цветковым координатам в одном цветовом пространстве предсказать вектор цветковых координат для того же излучения в другом пространстве.

Сформулируем задачу поиска наилучшего такого преобразования как задачу минимизации эмпирического риска. Пусть цветовые пространства  $S_1$  и  $S_2$  задаются спектральными чувствительностями  $\vec{\psi}(\lambda)$  и  $\vec{\chi}(\lambda)$  соответственно. Пусть  $R(\vec{a}, \vec{b})$  — некоторая функция потерь, принимающая наименьшее значение при  $\vec{a} = \vec{b}$ . Нужно найти преобразование  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  минимизирующее матожидание

$$\mathbb{E}_{F \sim p(F)} R \left( \int F(\lambda) \vec{\psi}(\lambda) d\lambda, T \left[ \int F(\lambda) \vec{\chi}(\lambda) d\lambda \right] \right), \tag{4}$$

где  $p(F)$  — распределение спектров в реальных сценах.

Важный случай задачи — переход из пространства RGB камеры в пространство стандартного наблюдателя CIE XYZ (цветовое пространство человека без зрительных аномалий). Функции спектральной чувствительности XYZ показаны на рис. 2.

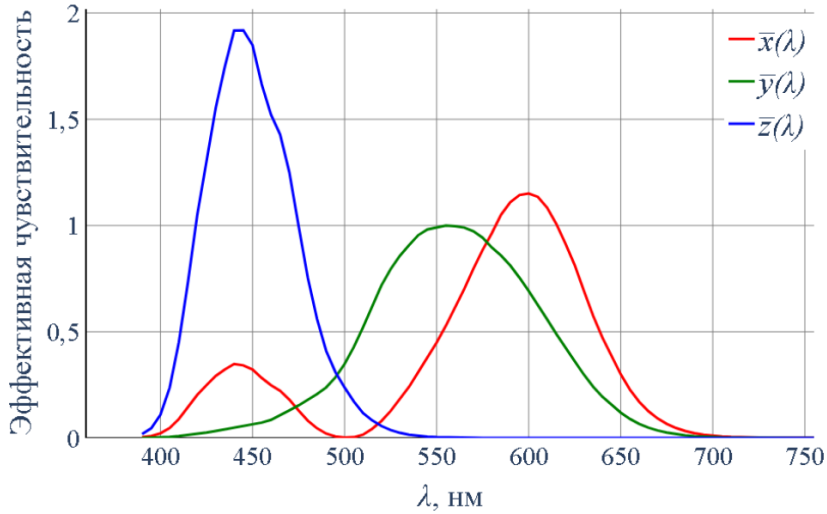


Рис. 2. Функции спектральной чувствительности XYZ, рисунок из учебного пособия [19].

Поскольку часто цветовой переход совершается не для отдельно взятого цвета, а для целого изображения (например, в рамках ISP), полезно ввести постановку задачи CST для изображения. Пусть  $I, J$  — изображения,  $R(I, J)$  — некоторая функция потерь (мера различия изображений), принимающая наименьшее значение при  $I = J$ . Нужно найти преобразование изображений  $T$ , минимизирующее матожидание

$$\mathbb{E}_{F \sim p(F)} R \left( \int F(\lambda, x, y) \vec{\psi}(\lambda) d\lambda, T \left[ \int F(\lambda, x, y) \vec{\chi}(\lambda) d\lambda \right] \right), \tag{5}$$

где  $p(F)$  — распределение спектральных изображений.

Обычно в качестве  $R(I, J)$  выбирается квадрат  $L_2$ -нормы:

$$T = \operatorname{argmin}_T \sum_{x,y} \|I(x, y) - T(J(x, y))\|_2^2. \tag{6}$$

### 3. НАБОР ДАННЫХ И МЕТРИКИ КАЧЕСТВА

Чтобы обучить нейросетевую модель CST и оценить качество полученного решения, необходим набор данных из пар изображений  $\{(I_i, I'_i)\}_{i=1}^N$ , где  $I_i$  — изображение в цветовом про-

пространстве  $S_1$ , а  $I'_i$  — изображение той же сцены в цветовом пространстве  $S_2$ . Простейший и самый точный способ получить такие данные — синтезировать их, используя набор гиперспектральных фотографий, содержащих в каждом пикселе значения спектральной плотности излучения для достаточно большого множества длин волн.

Чтобы получить из такой фотографии данные для обучения, можно симулировать регистрацию спектра сенсором численным интегрированием согласно формуле (2). Так как в реальности данные представляют из себя дискретный набор отсчетов вдоль длины волны, технически можно переписать модель в матричном виде:

$$I = FX^T, \quad (7)$$

где  $F$  — тензор размером  $H \times W \times C$ ,  $C$  — количество спектральных отсчетов,  $X$  — матрица чувствительностей  $3 \times C$ . Осуществив проекцию в пространство  $S_1$  и пространство  $S_2$ , мы получаем требуемую пару фотографий.

Отметим, что есть и другие способы получения необходимых для обучения и тестирования пар изображений. Так, можно использовать сопоставленные пары реальных изображений с двух различных фотоаппаратов. Однако между двумя фотографиями обязательно будет пространственная или временная разница, за счет чего сцены могут отличаться, и даже при хорошем сопоставлении изображений могут оставаться различия, делающие попиксельную оценку качества неточной. При генерации же из гиперспектральных фотографий таких эффектов не возникает, и итоговое влияние неидеальности оптической системы минимально.

Кроме наборов собственно гиперспектральных фотографий, существуют также наборы изображений, содержащих отражательные способности объектов в сцене,  $\Phi(\lambda, x, y)$  в формуле (3) после дискретизации.

Для обучения нейронной сети на преобразование изображения необходимо много данных, поэтому был выбран набор гиперспектральных данных KAUST [10]. Он содержит отражательные способности для 409 сцен, включая естественные и искусственные, в помещении и на открытом воздухе. Изображения были получены на камеру Specim IQ, отражательные способности восстановлены по белой цветовой мишени, находящейся в кадре (обычно около края). Спектральное разрешение: 34 канала для длин волн от 400 до 730 нм (шаг 10 нм). Пространственное разрешение  $512 \times 512$ . Чтобы в кадр не попадала белая цветная мишень (в большем числе сцен), фотографии были обрезаны до размера  $448 \times 448$ .

Часть изображений содержали источник освещения или были засвечены, что делает линейную модель формирования изображений неприменимой к ним. Всего в выбранном наборе данных обнаружилось 7 таких изображений. Для генерации изображений в разных цветовых пространствах они не использовались. Из оставшихся 402 изображений 80 были выбраны для тестовой выборки, остальные 322 — для обучающей.

Чтобы представить многообразие различных условия освещения, были выбраны 29 стандартных источников освещения, и спектр каждого умножался на каждое изображение в KAUST, увеличивая тем самым общий размер выборки до  $402 \times 29 = 11658$  изображений.

Выбранные источники включают излучение абсолютно черного тела (близкое к свету лампы накаливания) A18, A24 и A28, дневной свет D50, D55, D65, D75 и D95, 12 люминесцентных ламп FL1—FL12, светодиодные лампы LED-B1—B5, LED-BH1, LED-RGB1, LED-V1 и LED-V2.

После получения спектральных изображений, они были спроецированы в цветовые пространства XYZ и RGB камеры Nikon D700 согласно введенной линейной модели формирования изображения и нормированы для приведения значений к диапазону  $[0; 1]$ .

Для оценки качества CST кроме набора данных необходимо определить набор релевантных метрик. К ним мы отнесли стандартные метрики  $L_1$ ,  $L_2$  и более удобную для интерпретации метрику RMSE — корень из среднеквадратического отклонения. Также полезно рассмотреть



Рис. 3. Иллюстрация содержимого набора данных KAUST из статьи [10].

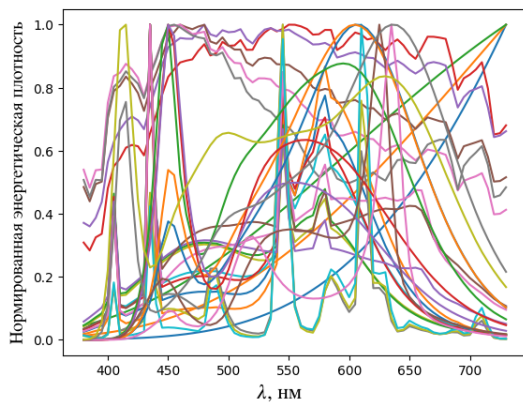


Рис. 4. Спектры использованных источников, нормированные на максимум.

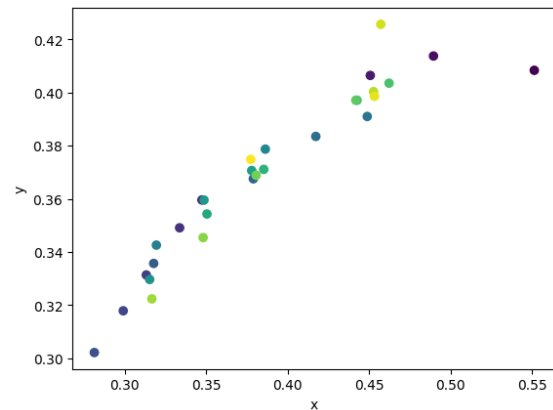


Рис. 5. Координаты использованных источников в плоскости цветности:  
 $x = X/(X + Y + Z)$ ,  $y = Y/(X + Y + Z)$ .

SSIM (Structural Similarity Measure, индекс структурного сходства) [17] — метрику, основанную на особенностях зрительного восприятия человека и учитывающую различия в яркости, контрастности и пространственной структуре изображений. В то время как для классических поточечных методов цветокоррекции SSIM не очень актуальна, в нашем случае использование метрики, отражающей различие в пространственной структуре, необходимо, поскольку сверточная нейросеть может изменять структуру изображения.

Отдельно выделим цветовые метрики, к которым мы относим угловое расстояние, косинусное расстояние, формулу CIEDE2000 [11],  $L_2$  и угловое расстояние в ProLab [6]. Пусть  $I$  и  $J$  — истинное и предсказанное изображения одинакового размера  $H \times W$ . Угловое расстояние определяется как

$$\text{Angle} = \frac{1}{HW} \sum_{i,j} \frac{180^\circ}{\pi} \arccos \frac{\langle J_{ij}, I_{ij} \rangle}{\|J_{ij}\|_2 \|I_{ij}\|_2}.$$

Поскольку угловое расстояние инвариантно к умножению векторов цветовых координат на скаляр, оно характеризует различие в цветности, что особенно актуально для задачи цветового перехода. Другой вариант угловой меры, сложнее интерпретируемый, но проще для вычисления — косинусное расстояние:

$$\rho_{\cos} = 1 - \frac{1}{HW} \sum_{i,j} \frac{\langle J_{ij}, I_{ij} \rangle}{\|J_{ij}\|_2 \|I_{ij}\|_2}.$$

Более точную количественную характеристику различия между цветами в соответствии с восприятием человека дают *формулы цветового различия*. Самой точной из них является CIEDE2000 ( $\Delta E_{00}$ ). Эта метрика учитывает адаптацию зрительной системы человека к освещению и потому зависит не только от цветовых координат сравниваемых цветов, но и от цветности источника освещения. Пространство ProLab — это пространство, связанное с XYZ проективным преобразованием, евклидово расстояние ( $L_2$ ) в котором моделирует ощущение цветового различия человека. Преобразование было подобрано так, чтобы  $L_2$ -расстояние в новых координатах как можно лучше приближало CIEDE2000, а также чтобы  $\vec{0}$  (черный цвет) переходил в  $\vec{0}$ .  $L_2$  в ProLab вычисляется по значительно более простой формуле, а проективная зависимость от координат в XYZ и сохранение нуля дают полезное свойство: лучи, выходящие из нуля переходят в лучи, выходящие из нуля. Таким образом, в ProLab на лучах, выходящих из нуля лежат векторы, соответствующие цветам с разной яркостью, но одинаковой цветностью. В силу вышесказанного, угловое расстояние в ProLab характеризует различие в цветности, причем более согласованно с человеческим восприятием, в отличие от углового расстояния в XYZ.

#### 4. МЕТОД

В качестве основы была взята модель LW-ISP [2], предложенная как полностью нейросетевой конвейер формирования изображений. Среди качественных нейросетевых ISP эта модель одна из наиболее легковесных.

LW-ISP имеет U-Net-подобную архитектуру с использованием блоков внимания в первой половине сети и специальных блоков увеличения разрешения во второй половине (см. рис. 6). В модели применяется механизм внимания отдельно по каналам и по пространственным измерениям, результаты конкатенируются. Блок увеличения разрешения использует операцию Pixel Shuffle вместо обратной свёртки, а также карты признаков с соответствующего уровня первой половины сети, полученные вычитанием результатов «прореженной» (dilated) и обычной свёрток. На вход нейросети подаётся четырёхканальное изображение — по одному каналу на красные и синие пиксели в фильтре Байера и два канала на зелёные. Так как в результате ISP нужно получить изображение исходного размера, операций увеличения разрешения в модели на одну больше, чем операций уменьшения разрешения.

Для обучения LW-ISP авторы предложили механизм гетерогенной дистилляции: модель-учитель с похожей архитектурой обучается решать более простую, но близкую задачу, после чего выходы её скрытых слоёв используются для облегчения обучения модели-ученика.

В этой работе архитектура и алгоритм обучения LW-ISP были несколько модифицированы для адаптации к решаемой задаче и ускорения вычислений.

Во-первых, для упрощения процесса обучения было решено не использовать дистилляцию. Во-вторых, для сжатия модели свертки были заменены на глубинно-разделяемые (тип свертки, использованный в архитектуре MobileNet), как было предложено в работе [20]. Поскольку оригинальная нейросеть увеличивает размер изображения в 2 раза, были убраны последние слои, увеличивающие разрешение. Количество входных каналов заменено с 4 на 3. К функции

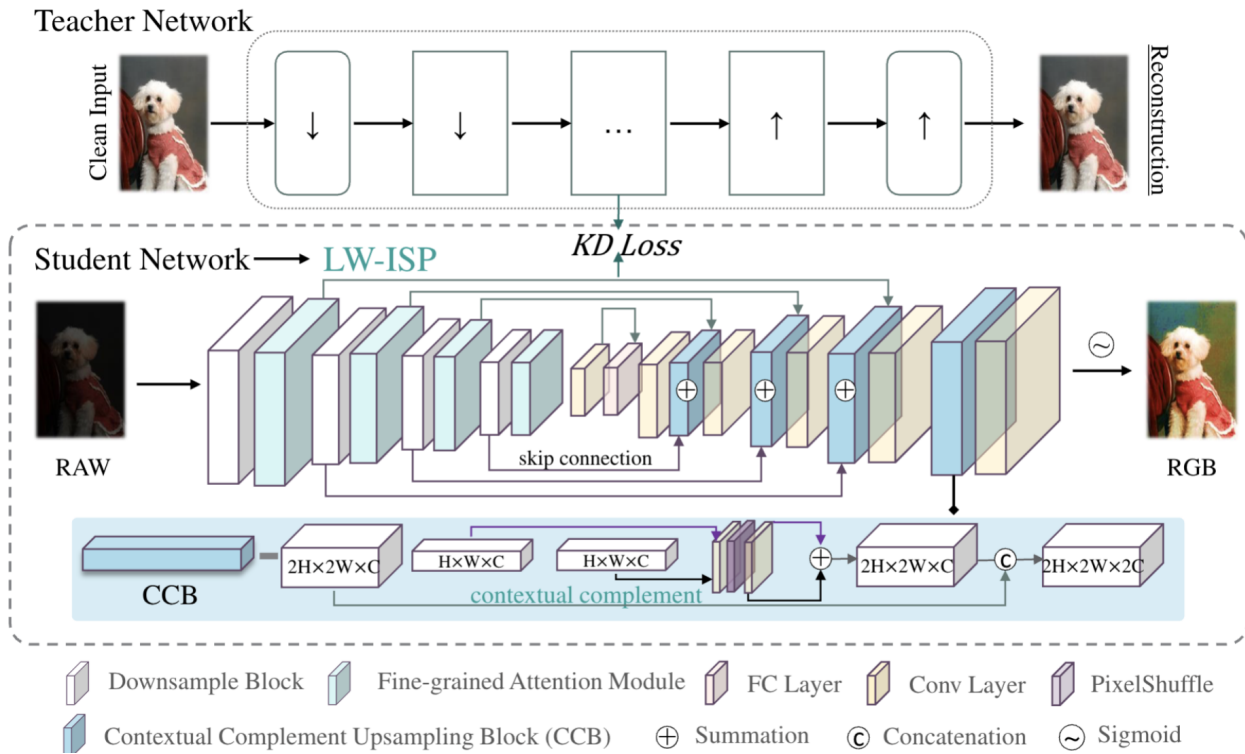


Рис. 6. Схема архитектуры LW-ISP из оригинальной статьи [2].

потерь прибавлено косинусное расстояние, чтобы штрафовать нейросеть за неудовлетворительно предсказанную цветность. Итоговая функция потерь имеет вид

$$\mathcal{L} = \mathcal{L}_r + \alpha \cdot \mathcal{L}_s + \beta \cdot \mathcal{L}_c, \quad (8)$$

где

$$\begin{aligned} \mathcal{L}_r &= L_1(I_2, T(I_1)) \\ \mathcal{L}_s &= 1 - SSIM(I_2, T(I_1)) \\ \mathcal{L}_d &= \rho_{\cos}(I_2, T(I_1)), \end{aligned}$$

$\alpha = 0, 2$ ;  $\beta = 0, 4$ .

Также убрана сигмоида на выходе нейросети. Первые эксперименты показали, что сеть показывает себя хуже с логистической функцией активации в конце, чем без активации. Использование же активации ReLU вовсе приводит к расходимости процесса обучения.

## 5. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

В ходе работы было проведено множество экспериментов по обучению различных моделей с разными гиперпараметрами и разными вариантами постановки задачи. В этой главе представлены основные результаты и параметры наиболее успешных запусков.

Предложенная система оценки качества была реализована с использованием языка Python и библиотеки PyTorch. Были реализованы основные классические методы, поточечные нейросетевые и модификация LW-ISP.

Для нейросетевых методов использовалась одна та же функция потерь (8), которая минимизировалась оптимизатором Adam. Для поточечных нейросетевых методов использовался шаг  $10^{-4}$ , для модификации LW-ISP —  $5 \cdot 10^{-4}$ , косинусное расписание шага, а также аугментация в виде случайного поворота изображения на  $90^\circ$ ,  $180^\circ$  или  $270^\circ$  и случайного отражения.



## 5.1. Тестирование методов цветового перехода

Основные классические методы и нейросетевые методы были обучены осуществлять переход из пространства камеры Nikon D700 в XYZ на сгенерированном наборе данных, затем оценены по всем рассмотренным метрикам. Итоговые оценки качества цветового перехода по результатам экспериментов представлены в таблице 1. Метрики  $L_1$ , RMSE, SSIM позволяют оценить степень изменения семантики сцены.  $\Delta E_{00}$ ,  $\Delta E_{\text{ProLab}}$ ,  $\text{Angle}_{\text{ProLab}}$  в свою очередь демонстрируют качество работы моделей с точки зрения цветности. Все метрики в таблицах показывают улучшение при уменьшении значения.

Так как неросетевые модели не обладают однородностью порядка 1, важным этапом валидирования качества решения является проверка устойчивости алгоритма к изменению яркости. Для этого были сравнены метрики качества цветопередачи при умножении изображения на определенные константы. Дополнительно были проведены эксперименты, где нейронной сети во время обучения в качестве аугментации подавались изображения с измененной яркостью. Результаты сравнения представлены в таблицах 2 и 3.

**Таблица 1.** Сравнение методов цветового перехода. Для каждой метрики жирным шрифтом выделено лучшее значение, подчеркнутым — следующее за ним.

«Лин. + среднее» — линейная модель с добавлением усредненного по всей сцене цвета как дополнительных 3 признаков. «FC» — полносвязная нейросеть, действующая поточечно, в скобках указано количество нейронов в скрытых слоях. «LW-ISP-Aug» — так же модифицированная модель LW-ISP с использованием яркостной аугментации.

Модель	$L_1$	RMSE	SSIM	Angle	$\Delta E_{00}$	$\Delta E_{\text{ProLab}}$	$\text{Angle}_{\text{ProLab}}$
Лин. модель	0,0043	0,0077	0,9961	1,271	2,428	1,323	1,703
Лин. + среднее	0,0042	0,0076	0,9963	1,234	2,384	1,295	1,697
PCC-2	0,0043	0,0076	0,9958	1,277	2,556	1,394	1,835
PCC-3	0,0042	0,0074	0,9964	1,210	2,411	1,318	1,703
RPCC-2	0,0043	0,0077	0,9965	1,207	2,339	1,279	1,618
RPCC-3	0,0043	0,0077	0,9964	1,231	2,353	1,291	1,643
RPCC-4	0,0043	0,0077	0,9965	1,222	2,342	1,286	1,635
SRCC-2	0,0043	0,0077	0,9962	1,210	2,352	1,285	1,627
SRCC-3	0,0042	0,0075	0,9966	1,192	2,293	1,257	1,598
FC (32)	0,0047	0,0090	0,9961	1,257	2,508	1,403	1,774
FC (64)	0,0046	0,0086	0,9961	1,252	2,500	1,391	1,750
FC (21, 77, 21)	0,0046	0,0087	0,9962	1,345	2,561	1,466	1,908
FC (79, 36)	0,0044	0,0081	0,9962	1,318	2,546	1,405	1,871
LW-ISP	<b>0,0024</b>	<b>0,0049</b>	<b>0,9985</b>	<b>0,807</b>	<b>1,393</b>	<b>0,763</b>	<b>1,043</b>
LW-ISP-Aug	<u>0,0027</u>	<u>0,0060</u>	<u>0,9980</u>	<u>0,891</u>	<u>1,579</u>	<u>0,881</u>	<u>1,194</u>

**Таблица 2.** Среднее значение метрики CIEDE2000 на задаче CST в зависимости от мультипликатора яркости  $\alpha$ .

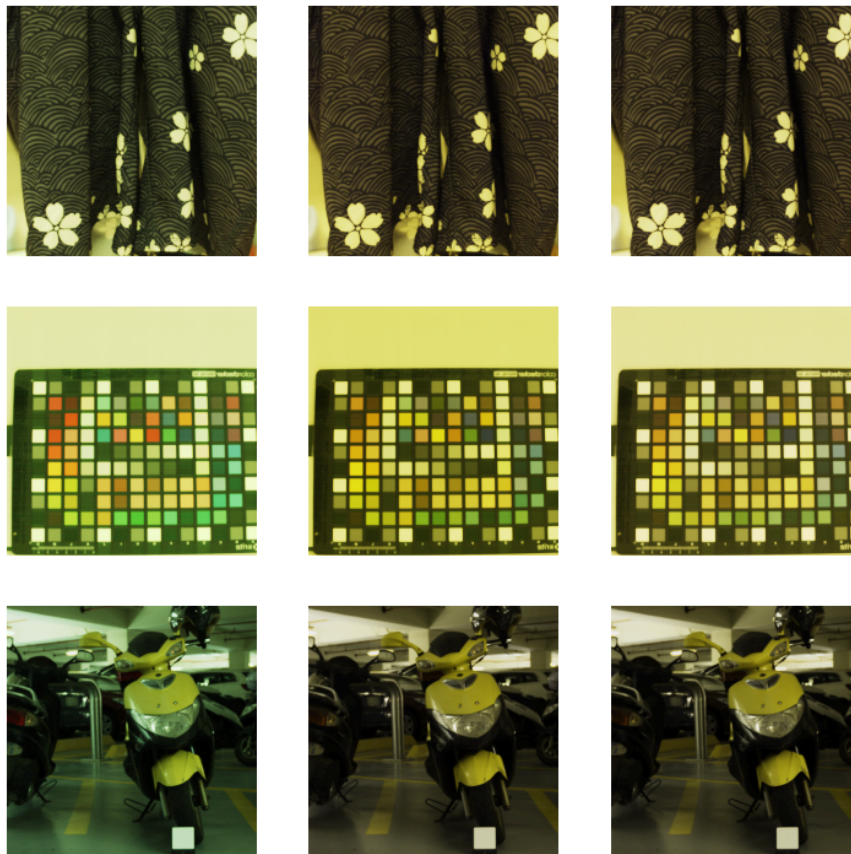
$\alpha$	0,2	0,5	1,0	2,0	5,0
Лин. модель	2,43	2,43	2,43	2,43	2,43
SRCC-3	2,29	2,29	2,29	2,29	2,29
LW-ISP	2,67	<b>1,69</b>	<b>1,39</b>	1,79	2,65
LW-ISP-Aug	<b>2,07</b>	<u>1,73</u>	<u>1,58</u>	<b>1,48</b>	<b>1,45</b>

**Таблица 3.** Среднее значение угловой ошибки в ProLab на задаче CST в зависимости от мультипликатора яркости  $\alpha$ .

$\alpha$	0,2	0,5	1,0	2,0	5,0
Лин. модель	1,70	1,70	1,70	1,70	1,70
SRCC-3	<b>1,60</b>	1,60	1,60	1,60	1,60
LW-ISP	2,09	<b>1,29</b>	<b>1,04</b>	<u>1,19</u>	1,67
LW-ISP-Aug	<u>1,65</u>	<u>1,35</u>	<u>1,19</u>	<b>1,09</b>	<b>1,05</b>

Как видно из таблицы, предложенная нейросетевая модель превзошла другие методы по всем рассматриваемым метрикам, оказавшись на 39% лучше SRCC по CIEDE2000 и на 35% лучше SRCC по угловому расстоянию в пространстве ProLab. Пример ее работы представлен на рис. 7.

При анализе устойчивости работы алгоритма к изменению яркости видно, что алгоритм заметно зависит от яркости изображений. Обучение модели с использованием яркостной аугментации не нивелирует эффект полностью, хотя и уменьшает разброс качества. Тем не менее, даже при сильных изменениях яркости нейросетевая модель показывает качество лучше, чем поточечные методы.



**Рис. 7.** Пример работы предложенного нейросетевого метода на трех изображениях. Слева визуализированы изображения в пространстве камеры, посередине — в пространстве XYZ, справа — результат работы модели.

### 5.2. Тестирование на совместной задаче цветового перехода с цветовой константностью

Предложенную систему с небольшой модификацией можно также использовать для оценки качества в смежной задаче цветового перехода совместно с цветовой константностью. Для этого будем брать для обучения и тестирования пары изображений, где в первом симулировано произвольное освещение, а во втором — стандартное D65.

Для такой задачи были проведены аналогичные эксперименты, полученные результаты представлены в таблице 4.

Можно увидеть, что в этой задаче методы на основе LW-ISP значительно превосходят остальные модели. Это ожидаемо, так как для цветовой константности нужно иметь информацию об источнике освещения, которую невозможно извлечь из отдельного пикселя без учета контекста.

**Таблица 4.** Сравнение рассмотренных моделей в применении к задаче цветового перехода совместно с цветовой константностью. Для каждой метрики жирным шрифтом выделено лучшее значение, подчеркнутым — следующее за ним.

Модель	$L_1$	RMSE	SSIM	Angle	$\Delta E_{00}$	$\Delta E_{\text{ProLab}}$	Angle <sub>ProLab</sub>
Лин. модель	0,0141	0,0300	0,9702	7,121	6,902	4,365	6,134
Лин. + среднее	0,0136	0,0290	0,9685	7,715	6,836	4,239	6,682
PCC-2	0,0125	0,0265	0,9748	6,867	6,438	4,079	5,867
PCC-3	0,0121	0,0253	0,9764	6,657	6,307	3,983	5,732
RPCC-2	0,0127	0,0272	0,9734	6,841	6,436	4,088	5,828
RPCC-3	0,0126	0,0269	0,9734	6,812	6,372	4,043	5,768
RPCC-4	0,0125	0,0268	0,9733	6,798	6,337	4,017	5,739
SRCC-2	0,0129	0,0276	0,9733	6,825	6,385	4,095	5,805
SRCC-3	0,0121	0,0261	0,9739	6,625	6,134	3,876	5,584
FC (32)	0,0147	0,0334	0,9728	6,997	6,704	4,400	5,986
FC (64)	0,0146	0,0334	0,9727	7,002	6,645	4,389	6,043
FC (21, 77, 21)	0,0132	0,0280	0,9729	7,242	7,490	4,587	6,919
FC (79, 36)	0,0140	0,0300	0,9737	6,895	6,645	4,354	5,925
LW-ISP	<b>0,0041</b>	<b>0,0106</b>	<b>0,9973</b>	<b>1,405</b>	<b>1,640</b>	<b>1,392</b>	<b>1,388</b>
LW-ISP-Aug	<u>0,0053</u>	<u>0,0129</u>	<u>0,9944</u>	<u>2,313</u>	<u>2,739</u>	<u>1,662</u>	<u>2,237</u>

## 6. ОГРАНИЧЕНИЯ

Несмотря на показанный потенциал применения, данная работа обладает рядом ограничений, которые не влияют на полученный результат. А именно, отсутствуют:

- исследование разных архитектур нейронной сети;
- оценка вычислительной эффективности методов;
- абляционное исследование возможных функций ошибки;
- сравнение методов с семейством параметризуемых методов [5].

## 7. ЗАКЛЮЧЕНИЕ

Вычислительные эксперименты показали превосходство предложенной контекстно-зависимой модели над поточечными методами цветового перехода и линейной моделью с глобальным контекстом. Предлагаемый метод снижает цветовое различие CIEDE2000 на 39% и уменьшает угловое расстояние в пространстве ProLab на 35%. Это подтверждает важность учета контекста в данной задаче, причем в виде сложных признаков, выделяемых нейронной сетью и характеризующих семантику сцены. Показана способность нейросети на основе LW-ISP хорошо решать смежную задачу цветового перехода с цветовой константностью, в отличие от других методов, которые справились с ней сильно хуже. Предлагаемый метод снижает цветовое различие CIEDE2000 на 73% и уменьшает угловое расстояние в пространстве ProLab на 75%. Полученные результаты еще раз подтверждают перспективность использования нейронных сетей в ISP и его подзадачах, показывая способность нейросетей качественно обрабатывать цвет. Перспективным направлением дальнейших исследований является повышение устойчивости алгоритма к изменению яркости путем обучения отдельно цветности и интенсивности, а так же уменьшение размера нейронной сети для обеспечения возможности внедрения в мобильные устройства.

## СПИСОК ЛИТЕРАТУРЫ

1. Simone Bianco, Claudio Cusano, and Raimondo Schettini. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, 26(9):4347–4362, 2017.

2. Hongyang Chen and Kaisheng Ma. LW-ISP: A lightweight model with ISP and deep learning. *arXiv preprint arXiv:2210.03904*, 2022.
3. Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. Color correction using root-polynomial regression. *IEEE Transactions on Image Processing*, 24(5):1460–1470, 2015.
4. Xueyang Fu and Xiangyong Cao. Underwater image enhancement with global–local networks and compressed-histogram equalization. *Signal Processing: Image Communication*, 86:115892, 2020.
5. Hakki Can Karaimer and Michael S. Brown. Improving color reproduction accuracy on cameras. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6440–6449, 2018.
6. Ivan A Konovalenko, Anna A Smagina, Dmitry P Nikolaev, and Petr P Nikolaev. Prolab: a perceptually uniform projective color coordinate system. *IEEE Access*, 9:133023–133042, 2021.
7. Abdullah Kucuk, Graham Finlayson, Rafal Mantiuk, and Maliha Ashraf. Comparison of regression methods and neural networks for colour corrections. In *London Imaging Meeting*, volume 3, pages 74–79. Society for Imaging Science and Technology, 2022.
8. Abdullah Kucuk, Graham D Finlayson, Rafal Mantiuk, and Maliha Ashraf. An exposure invariant neural network for colour correction. In *Color and Imaging Conference*, volume 30, pages 176–181. Society for Imaging Science and Technology, 2022.
9. Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern recognition*, 98:107038, 2020.
10. Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2672–2681, 2021.
11. M Ronnier Luo, Guihua Cui, and Bryan Rigg. The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research & Application*, 26(5):340–350, 2001.
12. Lindsay MacDonald. Color space transformation using neural networks. In *Color and Imaging Conference*, volume 27, pages 153–158. Society for Imaging Science and Technology, 2019.
13. Lindsay MacDonald and Katarina Mayer. Camera colour correction using neural networks. In *London Imaging Meeting*, volume 2, pages 54–57. Society for Imaging Science and Technology, 2021.
14. Konstantin Vitalyevich Soshin, Dmitry Petrovich Nikolaev, Egor Ivanovich Ershov, and Mikhail Konstantinovich Tchobanov. Scalable rational color correction for an image, 2024. US Patent App. 18/750,729.
15. R Von Luther. Aus dem gebiet der farbreizmetrik. *Zeitschrift fur Technische Physik*, 12:540–558, 1927.
16. Yang Wang, Jing Zhang, Yang Cao, and Zengfu Wang. A deep cnn method for underwater image enhancement. In *2017 IEEE international conference on image processing (ICIP)*, pages 1382–1386. IEEE, 2017.
17. Zhou Wang. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
18. Hefeng Zhan, Songxin Shi, and Yang Huo. Computational colour constancy based on convolutional neural networks with a cross-level architecture. *IET Image Processing*, 13(8):1304–1313, 2019.
19. Дмитрий Петрович Николаев, Пётр Петрович Николаев, Сергей Александрович Гладилин, and Валентина Петровна Божкова. *Основы цветовой теории в техническом зрении*. Мир науки, 2021.
20. Александр Срибняк, Артем Паншин, and Егор Ершов. Сжатие нейросетевого пайплайна формирования изображений. In *Информационные технологии и системы 2024 : сборник трудов 48-й междисциплинарной школы-конференции ИППИ РАН, Воронеж, Москва, 16–20 сентября 2024 года*, pages 98–105. Москва: Институт проблем передачи информации им. А.А. Харкевича РАН, 2024.
21. Бернд Яне. *Цифровая обработка изображений*. Техносфера, 2007.

## Color Space Transform using Convolutional Neural Networks

A. S. Sribniak, A. V. Panshin, E. I. Ershov

This work focuses on the study of color space transform — a key stage in the image signal processing pipeline. Existing methods primarily perform pointwise image transformation without considering contextual information. We propose a novel color space transform method based on the LW-ISP convolutional neural network, capable of incorporating scene semantics. The proposed method demonstrates significant improvements over SotA color transform approaches: achieving a 39% reduction in CIEDE2000 color difference and a 35% reduction in angular distance in the ProLab space.

**KEYWORDS:** color space transform, color correction, convolutional neural networks.